The Cult of Causality, Pearl's priesthood has no clothes

Benjamin James

November 28, 2025

Science likes to pretend it grew out of myth. The lab coat, p-value, and randomized controlled trials are all presented as a decisive break from the priesthoods and oracles of earlier eras. But when you look closely at how causal claims are manufactured and sold, you see something far less flattering. You see a priesthood with better notation. You see ritualized procedures for turning uncertainty into pronouncement. You see in Judea Pearl's causal framework not just a technical achievement, but a ready-made liturgy for a culture that would rather sound certain than be structurally aligned with reality. The danger is not that Pearl made causal reasoning computable. The danger is that his formalism arrived in a scientific ecosystem already addicted to certainty, and it became another instrument for performing knowledge rather than interrogating it.

Pearl's diagrams and do-operators did something important; they took the hand-wavy talk of "A causes B" and gave it syntax. Variables became nodes, assumptions became arrows, and interventions became algebraic operators. Suddenly the fuzzy question, "What if we changed X?" could be rendered as a crisp expression. In the right hands, this is invaluable. However, the right hands are rare, and even when they exist, they are embedded in institutions that reward certainty above all else. So, the formal apparatus, which in principle is agnostic about its own adequacy, was repurposed as an authority device. A causal graph printed in a paper or flashed on a slide feels like a settled map of reality, not what it actually is, an authored fiction with equations. The trick is subtle but devastating. Internal precision is quietly passed off as external truth.

The mechanics are straightforward. Pearl's framework lets you write down a set of variables you care about, draw arrows corresponding to the causal relationships you believe exist, and then use the resulting structure to compute what would happen under hypothetical interventions. None of this discovers the variables you omitted. None of it guarantees the arrows are right. The entire edifice rests on the modeler's choices, discipline norms, and the

current fashion of what "everyone knows" must be causally relevant. The structure is self-authored and self-contained. What the math then does, exquisitely, is preserve consistency within that sealed world. The more polished the consistency, the easier it is to forget that the world it governs is one we manufactured in advance. Pearl's calculus ensures that if your assumptions were right, your inferences will be coherent. It cannot tell you when your assumptions are shallow, parochial, or structurally delusional.

An example from psychopharmacology shows how deep the problem runs. The standard causal diagram used in SSRI research is almost offensively simple. The drug D increases synaptic serotonin S, which in turn reduces depressive symptoms Y. Confounders such as baseline severity C and socioeconomic stress E are treated as adjustable covariates. The resulting directed acyclic graph (DAG), $C \to D, C \to Y, E \to Y, D \to S \to Y$, appears to satisfy the backdoor criterion. Pearl's calculus then gives a clean expression for the causal effect of administering the drug:

$$P(Y \mid do(D = d)) = \sum_{c,e} P(Y \mid D = d, C = c, E = e)P(C = c, E = e)$$

In this sealed world, the intervention "set serotonin to a higher level by setting $D = d$" is coherent. It yields a crisp causal estimate, the kind that populates meta-analyses and clinical guidelines.

But this sealed world is imaginary. The brain is not a DAG; it is a metastable, recursive system in which monoamine levels, receptor densities, stress physiology, learning history, and social context co-evolve. Serotonin changes dopamine; dopamine changes motivation; motivation changes behavior; behavior changes stress; stress modulates serotonin receptor expression. Even the supposedly "exogenous" drug D is not exogenous. Expectancy effects, therapeutic alliance, side-effect perception, and withdrawal dynamics all feed back into the very symptoms the model treats as downstream outcomes. A clean arrow from D to S and from S to Y is not a simplification. It is a hallucination made diagrammatic.

This is precisely how internal precision becomes mistaken for external truth. The model's static, acyclic structure makes the intervention look surgical and the effect identifiable, even though no real intervention on a psychopharmacologic agent behaves like a do-operator. The

causal estimate is correct only inside the world defined by the graph. A world without plasticity, feedback, subjective experience, or time. Yet, once exported to public discourse, the estimate is treated as a property of the brain itself rather than of the model that authored it. The authority arises not from alignment with reality, but from the fact that the formalism cannot reveal the ways it fails to track reality at all.

Add the institutional layer and this subtle limitation metastasizes into public dogma. Scientists are not free-floating minds; they are employees of universities, firms, agencies, and journals. Those organizations are not designed to tolerate ambiguity. They are designed to maximize funding, influence, policy relevance, and prestige. None of those currencies are served by leading with, "Our model is an incomplete caricature, extrapolating from limited data under questionable assumptions." So, the language hardens. "According to our causal model" becomes "the data shows," which the journalist truncates to "scientists say," which the politician recycles into "the science is setttled." By the time the claim reaches public discourse, the scaffolding of modeling assumptions has vanished. What remains is Pearl's elegant machinery, now treated as a black box that turns data into causal truth.

The machinery itself is not neutral. Pearl's models are structurally static, directed graphs with fixed nodes and fixed arrows. They encode the world as a frozen architecture. That is enough if you are modeling toy systems, where causal relationships are stable and well-defined. It is an insult to reality when applied to societies, economies, ecosystems, or minds. In living systems, causal structure is not a fixed wiring diagram; it is a field of shifting tendencies, contingent on context, time, and history. Feedback loops bloom and collapse. Interactions change. New variables become salient as old ones fade. To act as if one can capture this in a single acyclic graph and then treat that graph as the causal skeleton of the world is not just naïve, it is structurally incoherent. A static model in a dynamic environment leaks alignment by design; its internal consistency is purchased at the cost of tracking a moving target.

Then there is the intervention logic. The do-operator is a conceptual sledgehammer; "$do(X = x)$" means cut all incoming arrows into X and force it to a given value. This is mathematically convenient. It lets you reason about the effect of flipping a variable while

holding everything else fixed in a controlled way. But the world does not recognize this operator. You do not "set" obesity, or education, or social trust. You perturb multiple interdependent processes, most of which you only dimly understand. Real interventions ripple through structures, mutate constraints, and alter the very variables the model treats as exogenous. The neat trick of pretending you can surgically sever X from its causes and flip it at will is fine as a thought experiment. It is lethal when smuggled into policy as if the intervention is nothing more than "apply $do(X = x)$ and read off the effect on Y." It trains institutions to think in binary levers where only entangled nudges exist.

Counterfactuals, Pearl's other great selling point, fare no better when dragged into the messy world of reality. His twin-network construction assumes you can meaningfully compare "the world that happened" with "the world that could have happened if we had intervened." But again, this assumes that the model already encodes all relevant structural features and that the hypothetical intervention produces a reachable alternate trajectory. In practice, many counterfactuals manufactured this way are ontologically empty. They imagine worlds that no real system could ever have occupied, then attach probabilities to them as if they were legitimate neighborhoods in possibility space. The formalism does not care; blind algebra chugs along. Public discourse, once again, is presented with outputs stripped of their conditionality and told, for example, that "this policy saved 200,000 lives" or "this treatment would have reduced mortality by 30%." The fact that these numbers rest on counterfactual structures that might be impossible in any real-world trajectory is never part of the conversation.

The real obscenity is not that Pearl's models are limited. All models are. The obscenity is that the limits are systematically hidden from the public, and often from the practitioners themselves. Causal diagrams confer an aura of finality. They invite a kind of epistemic laziness. Once the graph is drawn and the effect is "identified," the causal question is treated as closed. In reality, identifiability is a low bar. It means, roughly, that you can express the quantity you want in terms of observable distributions under your modeling assumptions. It does not mean the quantity is stable under structural shifts, ethically relevant, or even the right causal question to ask. A model can be formally correct and practically useless, or

worse, practically harmful. The incentive structure in science privileges the formal property, “we identified this effect," over the structural one, “this effect is actually a meaningful handle on the system.”

Once this pattern is baked into the way scientists talk, journalists follow. They have neither the time nor the incentive to unspool the tower of assumptions behind every causal claim. So causal diagrams and do-operator results become a kind of epistemic weaponry in public discourse. “Follow the science” becomes an instruction to submit to whatever model currently dominates a given field, no matter how provisional its structural fit. Disagreement is framed not as a dispute over competing, imperfect models, but as an attack on science itself. The authority of the method is fused with the authority of specific instantiations of the method. Pearl’s machinery, which should have been a way to keep track of our uncertainty about causation, becomes instead a badge of infallibility.

The implications for democratic deliberation are dire. Causal models are increasingly deployed to justify policies with enormous distributional consequences, such as school closures, economic sanctions, climate interventions, and algorithmic decision systems. In each case, a small priesthood of experts selects variables, draws arrows, runs do-calculus, and produces a set of recommended actions. The models are often opaque even to other scientists, let alone to the public. Yet their outputs enter the discourse dressed as inevitabilities. “If we don’t do X, Y will happen.” The space of alternative models, alternative interventions, and alternative values is compressed out of existence. What remains is technocratic inevitabilism, the sense that causal science has already computed the answer, and the role of politics is simply to comply.

There is an additional, corrosive secondary effect. When these confident causal proclamations inevitably collide with the recalcitrant complexity of the real world, trust collapses. Policies justified as near-certainties fail, predictions miss, and interventions carry side effects the models did not anticipate because the models could not even represent the relevant dimensions. The public, having been trained to equate “science” with ironclad prediction, rightly senses betrayal. Instead of diagnosing the problem as institutional overreach and model hubris, discourse slides into crude anti-science reaction or tribal

alignment with whichever experts affirm one's prior beliefs. The space for honest acknowledgment of uncertainty shrinks further. The result is a feedback loop where overconfident causal models produce failed predictions; failed predictions produce distrust; distrust produces polarized demands for even more definitive scientific backing; and, so, the models are made even more assertive to maintain authority.

None of this is necessary. Pearl's framework could instead be used as it, structurally, ought to be, a language for making our assumptions explicit and for testing the internal consistency of our causal stories. His formalism is perfectly capable of being embedded in a culture that says, "Here is our best current attempt to formalize the causal structure we think we see; here are the ways it might be wrong; here are the regions of possibility space where we know we are blind." What forbids this is not the math but the institutional insistence on speaking with one voice and one level of confidence. Institutions want a single graph, a single estimated effect, a single recommendation. They are not built to traffic in moving distributions over models, to foreground option space, or to admit that under structural uncertainty the honest answer is typically, "We do not know yet, and our interventions will be partly exploratory."

A more honest causal practice would invert the current hierarchy of virtues. Instead of rewarding the cleanest narrative with the boldest claim, it would reward models that preserve option space, remain sensitive to structural shifts, and attach visible confidence bands not only to parameter estimates but to the graph itself. It would foreground the fact that any causal model is a snapshot under constraints, not the final word on how the world works. It would treat counterfactuals as fragile thought experiments, not as policy-grade predictions. And crucially, it would communicate these limitations directly to the public, not bury them in technical appendices.

That kind of practice would demand a different relationship between science and society. It would require a public capable of tolerating ambiguity, and institutions willing to admit ignorance without imploding. It would mean giving up the comforting hallucination that causality has been solved and returning to the harder work of tracking a world that is constantly escaping our graphs. This is not a retreat from rigor. It is rigor finally aimed where

it belongs, the boundary between our models and the world, rather than at polishing the interior consistency of systems we already know are incomplete.

As long as we pretend that a static graph with a few arrows and a do-operator can stand in for the tangled, evolving causality of real systems, we will continue to confuse formal neatness with understanding. We will continue to turn scaffolding into monuments, then act surprised when they collapse under stress. The blame will be misplaced on “anti-science sentiment,” “misinformation,” or anything except the one thing that actually needs to change, the way science talks about what it does and does not know.

The first step is simple. Stop selling causal models as certainties. Stop using Pearl’s machinery as a stamp of inevitability. Treat it, and present it, as what it is, a disciplined way of writing down guesses about how the world might work, useful precisely to the extent that we remember they are guesses. Until that happens, every new causal graph exported into public discourse is not just an inference. It is a fresh layer of varnish on a culture-wide performance of certainty that we can no longer afford.

In the end, the modern causal science is an unwitting priesthood not merely metaphorically, but structurally. A small cadre of experts selects the sacred variables, inscribes the arrows of doctrinal causation, and performs the ritualized intervention of the sacred do-operator, as if parting the veil of nature itself. The resulting pronouncements are delivered with the same ceremonious certainty that once accompanied auguries and oracles; dissent is treated less as theoretical challenge than as heresy against “science.” Pearl did not set out to create a clergy, but his formalism, so precise, elegant, and easily misused, now serves as the vestment through which institutions disguise their fear of uncertainty. Their robes may be mathematical rather than theological, but the function is the same. Present a façade of control over forces we barely comprehend, sanctify assumptions as truth, and reassure the public that someone, somewhere, truly knows how the world works.

References

James, B. (2025). From Causation to Coherence: Deconstructing Pearl and Reconstructing Reason. PhilPapers. https://philpapers.org/rec/JAMFCT